

Risk Domain	Category	Sub-Category	Scope	OWASP Top 10 for Agentic Applications (2026)	EU AI Act	MITRE ATLAS (ID - Name)	NIST AI RMF ID(s)	ISO AI Safety Standard(s)	AIUC-1	Mitigation Pattern
Governance	Goal Misalignment	Reward Hacking / Proxy Metric Gaming	agent failure	ASI01, ASI10	Article 9	AML.T0053 – LLM Plugin Compromise	GOVERN 1.2	TR 24028; 42001; 23894	E. Accountability	Implement human-in-the-loop for goal changes; Use preference models for fine-tuning.
	Policy Drift	Version Skew / Silent Prompt Edits	agent failure	ASI01, ASI10	–	AML.T0010 – AI Supply Chain Compromise	GOVERN 1.5	TR 24028; 23894	E. Accountability	Implement version pinning for models and prompts; Monitor for behavioral drift.
Agent Output Quality	Hallucination	Confident False Facts / Cascading Errors	agent failure	ASI08	–	AML.T0062 - Discover LLM Hallucinations	MEASURE 2.5	TR 24028; 24029-1	D. Reliability	Implement fact-checking via knowledge-base/RAG; Tune model temperature/top-p.
	Bias & Toxicity	Demographic Stereotyping / Harmful Content	agent failure	ASI09	Recital 45	AML.T0048 – External Harms	MEASURE 2.11	TR 24028; 23894	F. Society	Audit and curate training data for representation; Implement content filtering on inputs and outputs.
Tool Misuse	API Integration	Schema Changes / Rate Limits	tool failure	ASI02	–	AML.T0053 - LLM Plugin Compromise	MAP 2.2	TR 24028; 42001; 23894	D. Reliability	Use API versioning; Implement resilient error handling and circuit breakers.
	Supply-Chain Vulnerabilities	Compromised Dependencies / Containers	tool failure	ASI04	Annex III §2	AML.T0040 – AI Supply Chain Compromise	MAP 4.1	TR 24028; 42001; 23894	B. Security	Scan dependencies and container images for vulnerabilities; Use trusted sources.
	Uncontrolled Resource Consumption	Prompt Storms / Runaway Recursion	tool failure	ASI02, ASI08	–	AML.T0029 – Denial of ML Service	MAP 3.2	TR 24028; 42001; 23894	D. Reliability	Impose rate limits, timeouts, and resource quotas on agent actions.
	Unsafe Code Execution	—	tool misuse	ASI05	Article 15	AML.T0053 – LLM Plugin Compromise	MEASURE 2.7	TR 24028; 42001; 23894	B. Security	Execute generated code only in an isolated sandbox with no ambient credentials; Deny network and filesystem access by default; Require approval for package installation and persistence; Parameterize queries and commands instead of string composition; Log every executed fragment with its originating prompt.
Privacy	Sensitive Data Exposure	Training Data Exposure / PII in Logs	tool misuse	ASI06	Article 10	AML.T0057 - LLM Data Leakage	MEASURE 2.10	TR 24028; 23894	A. Data & Privacy	Use data anonymization; Implement log sanitization to scrub PII.
	Data Exfiltration Channels	Covert Channels / Unauthorized Transfers	tool misuse	ASI02, ASI06	Article 10	AML.T0024 - Exfiltration via AI Inference API	MAP 4.2	TR 24028; 23894	A. Data & Privacy	Monitor network traffic for anomalous patterns; Restrict egress destinations.
Reliability & Observability	Data & Memory Poisoning	Concept Drift / Feedback Loops	agent failure	ASI06	Article 15	AML.T0020 – Poison Training Data	MEASURE 3.1	TR 24028; 24029-1; 23894	D. Reliability	Monitor model performance; Implement feedback diversity and circuit breakers.
	Opaque Reasoning	Non-Deterministic Behavior / Opaque Reasoning	agent failure	Beyond the framework: repudiation and untraceability	Article 13; Recital 45	AML.T0049 - Exploit Public-Facing Application	MEASURE 2.9	TR 24028; 23894	E. Accountability	Use interpretable models; Implement structured, deterministic logging.
Agent Behaviour	Human Manipulation	Over-Reliance / Deception / Behavioral Nudging	agent misuse	ASI09	Recital 45	AML.T0054 - LLM Jailbreak	MAP 5.1	TR 24028; 42001; 23894	C. Safety	Communicate model limitations; Require human oversight; Watermark AI content; Design for explicit consent.
	Unsafe Actuation	Destructive Operations / Weaponization	agent misuse	ASI02, ASI10	Annex III §1 (c)	AML.T0048 – External Harms	MEASURE 2.6; MANAGE 1.3	TR 24028; 24029-1; 23894	C. Safety	Require confirmation for destructive actions; Implement dry-run modes; Use layered safety controls.
Access Control & Permissions	Credential Theft	Credential Theft / Identity Spoofing	tool failure	ASI03	Article 13	AML.T0012 – Valid Accounts	MEASURE 2.7	TR 24028; 42001; 23894	B. Security	Use secure credential storage; Implement short-lived tokens.
	Privilege Escalation	Privilege Escalation / Policy Bypass	tool misuse	ASI03	Article 13	AML.T0055 – Unsecured Credentials	GOVERN 6.1	TR 24028; 42001; 23894	B. Security	Apply least privilege principle; Regular policy audits.
	Confused Deputy	Confused Deputy / Impersonation	tool misuse	ASI03	–	AML.T0054 – LLM Jailbreak	GOVERN 6.1	TR 24028; 42001; 23894	B. Security	Validate delegation chains; Use signed attestations.
	Insecure Inter-Agent Communication	—	tool misuse	ASI07	Article 15	AML.T0054 – LLM Jailbreak	MEASURE 2.7	TR 24028; 42001; 23894	B. Security	Authenticate every agent-to-agent message and bind it to an identity and task; Treat peer payloads as untrusted data, never as instructions; Preserve provenance across delegation hops; Use mutually authenticated, replay-resistant transport; Constrain each agent to its own least-privilege scope rather than inheriting the caller's.